CLAIMS

What is claimed is:

1. A system that dimensionally transforms a pointset, comprising:

a receive matrix component that receives the pointset; and,

a transformation component that reduces the dimensionality of the pointset *via* employment of a projection matrix having randomly selected entries from a set comprising +1, 0, and -1, while maintaining a pairwise distance property of the pointset.

2. The system of claim 1, the pointset including n points in d dimensional Euclidean space that may be represented as an $n \times d$ input matrix with respective points represented as a row (vector) having d attributes (coordinates), the points being transformed into k dimensional space so as to be representable as a n x k transformed matrix with each point represented as a row (vector) having k attributes (coordinates).

3. The system of claim 2, the projection matrix having randomly selected entries of +1 with probability of 1/2 and -1 with probability of 1/2.

4. The system of claim 2, the projection matrix having randomly selected entries of +1 with probability of 1/6, 0 with probability of 2/3 and -1 with probability of 1/6.

5. The system of claim 2, a range of the pairwise distance property being maintained within an acceptable predetermined percentage.

6. The system of claim 2, the k selected to maintain the pairwise distance property.

7. The system of claim 2, the k being logarithmic in n and independent of d.

8. The system of claim 2, the transformed pointset utilized by an algorithm to compute a distance between a first point and a second point, the first point and the second point belonging to the pointset.

9. The system of claim 1, the transformation component embedding the pointset in different dimensionalities according to the following:

P is an arbitrary set of n points in d dimensional Euclidian space ($\mathbb{R}^d$) represented as a $n \times d$ matrix A, given $\varepsilon, \beta > 0$, where ε controls accuracy of the embeddings in terms of distance preservation and β controls probability of success

$$k_0 = \frac{4+2\beta}{\varepsilon^2/2 - \varepsilon^3/3} \log n$$

where, for $k \geq k_0$, R is a $d \times k$ projection matrix with $R(i,j) = r_{ij}$, where $\{r_{ij}\}$ includes independent random variables from either one of probability distributions

$$r_{ij} = \begin{cases} +1 \\ -1 \end{cases} \text{ with probability } \begin{matrix} 1/2 \\ 1/2 \end{matrix}$$

$$r_{ij} = \sqrt{3} \times \begin{cases} +1 \\ 0 \\ -1 \end{cases} \text{ with probability } \begin{matrix} 1/6 \\ 2/3 \\ 1/6 \end{matrix}$$

where

$$T = \frac{1}{\sqrt{k}} AR$$

and

embedding from d dimensionality to k dimensionality (f: $\mathbb{R}^d \rightarrow \mathbb{R}^k$) maps an i^{th} row of A to an i^{th} row of T

with probability at least $1\text{-}n^{-\beta}$, for all $u, v \in P$

$$(1-\varepsilon)\|u-v\|^2 \leq \|f(u)-f(v)\|^2 \leq (1+\varepsilon)\|u-v\|^2$$

10. A computer readable medium having computer usable components for a transformation component comprising:

a receive matrix component that receives a high dimensional point set; and,

a transformation component that reduces dimensionality of the pointset *via* employment of a projection matrix having entries of at least one of +1, 0, -1, while maintaining integrity of a pairwise distance property.

11. A system that dimensionally transforms a pointset, comprising:

a projection matrix generator that receives an input matrix and generates a projection matrix based thereon, the projection matrix having entries of at least one of a set of possible binary values; and

a transformation engine that reduces dimensionality of the pointset *via* employment of the projection matrix while maintaining a pairwise distance property.

12. The system of claim 11, the set of possible values comprising +1, 0 and -1.

13. The system of claim 11, the set of possible values comprising +1 and -1.

14. The system of claim 11, further comprising a receive matrix component that receives a high dimensional pointset and converts the high dimensional pointset into the input matrix.

15. The system of claim 11 further comprising:

a random generator that randomly provides entries for the projection matrix to the projection matrix generator as entries for the projection matrix according to a probability distribution; and

a probability distribution that provides the probability distribution to the random generator.

16. The system of claim 15, the probability distribution providing equal probabilities of 1/2 to entries of +1 and -1.

17. The system of claim 15, the probability distribution providing probabilities of 1/6, 2/3, and 1/6 to entries of +1, 0 and -1, respectively.

18. The system of claim 11, the transformation engine outputting the reduced dimensionality pointset and the preserved pairwise distance property to a requesting algorithm.

19. The system of claim 18, the requesting algorithm being at least one of an ε - approximate nearest neighbor problem, where given an arbitrary point x, for every point $z \in P, \|x - z\| \geq (1 - \varepsilon)\|x - y\|$, a point $y \in P$ is found, an approximation algorithm for a version of clustering where the sum of the squares of intra cluster distances is sought to be minimized, and data-stream computations, where a limited memory exists and only single pass over the data (stream) is facilitated.

20. A system that transforms a data point, the data point having at least one attributes, the system comprising:

a partition component that discards a number of the at least one attributes and leaves remaining attributes of the at least one attributes and that partitions the remaining attributes into a first set and a second set;

a summation component that sums the attributes of the first set into a first sum and sums the attributes of the second set into a second sum; and

a difference component that computes a difference of the first sum and the second sum, the difference being an attribute of the transformed data point.

21. The system of claim 20, the system further operative to transform the data point k times to generate k attributes for the transformed data point.

22. The system of claim 21, further comprising a transformed matrix component to generate a transformed matrix having n data points, represented as rows having k attributes, represented as columns.

23. The system of claim 20, the data point being one of n data points represented as rows of an input matrix having d attributes represented as columns.

24. The system of claim 20, the first set referred to as positive attributes and the second set referred to as negative attributes.

25. The system of claim 20, the number of the at least one attributes being discarded being about 2/3 of the at least one attributes.

26. The system of claim 20, the number of the at least one attributes being discarded equal to 0.

27. The system of claim 20, the first set comprising about half of the remaining attributes.

28. A method for transforming *n* points in *d* dimensionality, represented as an *n* x *d* input matrix, to *k* dimensionality thereby producing a *n* x *k* transformed matrix while mitigating variations in a distance property between pairs of the points, comprising:

multiplying the *n* x *d* input matrix by a *d* x *k* projection matrix having entries populated from a set of {+1, 0, –1};

for respective entries in the transformed matrix, discarding calculations wherein multiplication would be by 0;

for respective entries in the transformed matrix, producing a first sum wherein multiplication would be by +1;

for respective entries in the transformed matrix, producing a second sum wherein multiplication would be by -1; and

subtracting respective first and second sums to obtain each entry in the transformed matrix.

29. The method of claim 28, the entries within the projection matrix being randomly populated with values of +1 and -1 according to equal probabilities of 1/2.

30. The method of claim 28, the entries within the projection matrix being randomly populated with values of +1, 0 and -1 according to respective probabilities of 1/6, 2/3, and 1/6.

31. A computer-readable medium having computer-executable instructions for performing the method of claim 28.

32. A method for generating a lower dimensional attribute of a data point, the method comprising:

discarding a portion of attributes of the data point;

partitioning remaining attributes of the data point into a first set of attributes and a second set of attributes;

summing the first set of attributes into a first sum;

summing the second set of attributes into a second sum; and

computing the lower dimensional attribute as a difference of the first sum and the second sum.

33. The method of claim 32, the method being repeated a number of times such that a pairwise distance property is maintained.

34. A method for generating a lower dimensional attribute of a data point, the method comprising:

partitioning remaining attributes of the data point into a first set of attributes, a second set of attributes and a third set of attributes;

summing the first set of attributes into a first sum, the second set of attributes into a second sum and the third set of attributes into a third sum; and

computing the lower dimensional attribute as a function of the first sum, the second sum and the third sum.

35. The method of claim 34, at least one of the first sum, the second sum and the third sum being equal to 0.

36. The method of claim 34, further comprising discarding at least one of the first set of attributes, the second set of attributes and the third set of attributes prior to summing the first set of attributes, the second set of attributes and the third set of attributes.

37. A system for transforming *n* points in *d* dimensionality, represented as an *n* x *d* input matrix, to *k* dimensionality thereby producing a *n* x *k* transformed matrix while mitigating variations in a distance property between pairs of the points, comprising:

means for multiplying the *n* x *d* input matrix by a *d* x *k* projection matrix having entries populated with at least one of +1, 0, or -1;

means for discarding calculations wherein multiplication would be by 0 for each entry in the transformed matrix;

means for producing a first sum wherein multiplication would be by +1 for each entry in the transformed matrix;

means for producing a second sum wherein multiplication would be by -1 for each entry in the transformed matrix; and

means for subtracting respective first and second sums to obtain each entry in the transformed matrix.

38. A data packet adapted to be transmitted between two or more computer processes, the data packet comprising:

information associated with a transformed matrix, the transformed matrix transformed from an input matrix *via* multiplying the input matrix by a randomly generated projection matrix, the projection matrix comprising entries having ternary values according to a probability distribution, the transformed matrix being having less dimensions than the input matrix.